5-Ч

УДК 004.942

О.Р. Чертов, Д.Г. Павлов, В.В. Мальчиков, М.В. Александрова
Національний технічний університет України «Київський політехнічний інститут», м. Київ
chertov@i.ua

Виявлення аномальної поведінки користувача системи контекстної реклами

У статті розглядається ринок інтернет-реклами, зокрема контекстної, формалізується взаємодія об'єктів та суб'єктів процесу проведення рекламної кампанії в мережі. В результаті аналізу взаємозв'язків будуються моделі поведінки можливих зловмисників в системі та демонструється можливість застосування недиадного вейвлет-перетворення для визначення наявності аномальної поведінки користувачів.

Вступ

Інтернет все глибше проникає у різноманітні сфери діяльності суспільства. Ця тенденція не оминула й рекламний ринок. З кожним роком доля інтернет-реклами зростає як в Україні (рис. 1), так і в цілому в світі (рис. 2) [1-6]. Це зумовлено більш високою ефективністю, відносною дешевизною та легкістю проведення рекламної кампанії, порівняно із рекламою в традиційних засобах масової інформації.

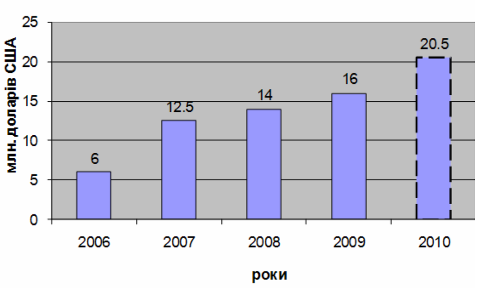


Рисунок 1 – Розвиток ринку інтернет-реклами в Україні

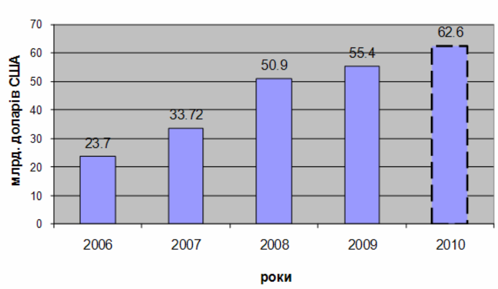


Рисунок 2 – Розвиток світового ринку інтернет-реклами

Найпопулярнішим різновидом реклами в мережі Інтернет є контекстна реклама. При її використанні рекламні оголошення демонструються на суміжних за тематикою сайтах або як результати пошуку за визначеними ключовими словами. Тематика контекстних рекламних посилань пов'язана із тематикою «навколишнього середовища». Завдяки цьому суттєво підвищується ймовірність зацікавленості користувача у продукті, який рекламується.

Оплата за розміщення оголошення контекстної реклами здійснюється залежно від обраної моделі рекламної кампанії: CPC (cost per click), CPM (cost per mille), CPA (cost per action), коли рекламодавець сплачує кожен клік, кожну тисячу показів або реєстрацію конверсії (деякої заздалегідь визначеної дії користувача).

Проте рекламні повідомлення в Інтернеті є вразливими до специфічного шахрайства: склікування та споказування. Склікування (click fraud) або споказування (impression fraud) – це різновид мережного шахрайства, коли імітується клік або показ оголошення [7]. Цілі проведення подібних заходів можуть бути різними: вичерпання

рекламного бюджету конкурента, отримання вищої позиції при показі оголошення зловмисника, власне збагачення (наприклад, власник сайта, на якому розміщене рекламне повідомлення, отримує деяку плату за кожен перехід за оголошенням).

На даний момент існують певні методи боротьби із мережним шахрайством: фільтрація всіх кліків в режимі реального часу (шаблони Google [8]), використання альтернативних моделей проведення рекламних кампаній, як-то: плата за дію (CPA), плата за відсоток показів (pay per percentage of impressions) [9]. Проте мережне шахрайство ще має значні масштаби – об’єм недійсних кліків оцінюється від 5% до 50% [10].

Метою даної роботи є аналіз взаємодії учасників процесу проведення рекламної кампанії в Інтернеті і побудова моделей поведінки можливих зловмисників. Отримані моделі можна використовувати для визначення наявності аномальної поведінки користувача системи контекстної реклами.

Постановка задачі

Під час проведення будь-якого нападу типова психологія поведінки зловмисника відповідає моделі інформаційної атаки (рис. 3).

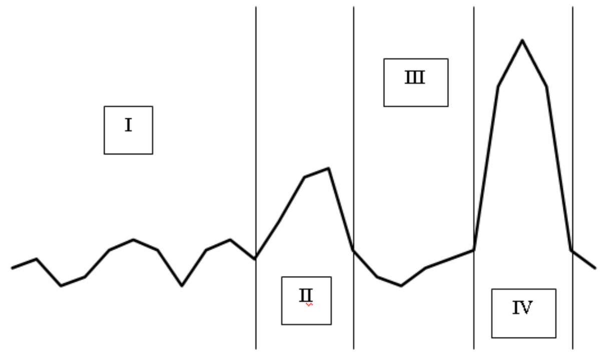


Рисунок 3 – Модель інформаційної атаки

Ця модель включає фази: «фоновий шум» (I), «проба» (II), «затишшя» (III), «атака» (IV), які у випадку мережного шахрайства відповідають природному рівню кліків по оголошенню (звичайна активність користувачів), першій спробі проведення нападу, вичікуванню результатів та склікуванню (споказуванню) оголошення.

Цілі та особливості проведення нападу і, відповідно, конкретний вигляд шаблону поведінки відрізняються залежно від того, до якої групи учасників належить зловмисник. Тому можна визначити моделі проведення подібного нападу кожною з можливих груп учасників і використовувати отримані результати для визначення наявності аномальних дій.

Моделювання взаємодії учасників процесу

В процесі створення та підтримки рекламної кампанії в Інтернеті беруть участь такі суб’єкти: рекламодавець, рекламне агентство, компанія, що надає послуги інтернет-реклами (рекламна компанія), власник сайта, що використовується як рекламний майданчик, та відвідувачі (люди, які переглядають рекламне повідомлення). На рис. 4 подано графічну ілюстрацію взаємодії учасників процесу.

Звичайні прямокутники зображають об’єкти процесу, закруглені – суб’єкти. Пунктирними фігурами зображені рекламні майданчики, а точковими лініями показано напрямок дії відвідувачів (тобто мету, з якою ними був здійснений перехід по оголошенню).

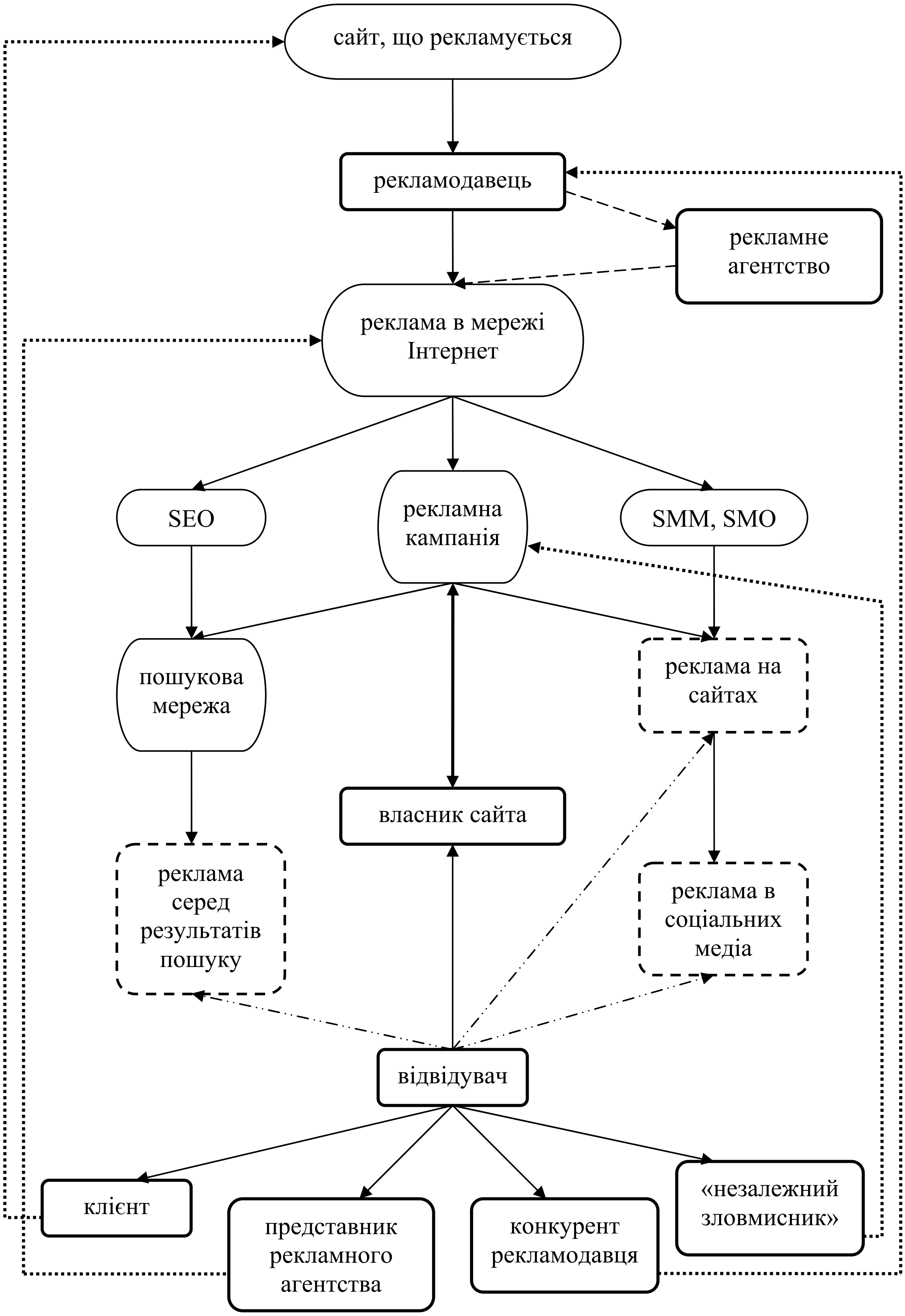


Рисунок 4 – Взаємодія учасників процесу

Процес створення будь-якої рекламної кампанії починається з вибору сайта, що буде рекламуватися, адже кожне рекламне повідомлення є посиланням на відповідний сайт. Після цього рекламодавець може запустити рекламу цього сайта за допомогою таких компаній, як Google Adwords, Yandex Direct, Begun, які розміщують рекламні посилання серед результатів пошуку та на різноманітних сайтах. Рекламодавець платить за розміщення оголошень залежно від обраної моделі рекламної кампанії: CPC, CPM, CPA (для контекстної реклами).

Як реклама в мережі Інтернет також використовуються різні види оптимізації. Перш за все це SEO (search engine optimization) або пошукова оптимізація. Задачею SEO є виведення сайта на перші позиції серед результатів пошуку. Для цього використовуються як легальні методи (наповнення сайта більш релевантним змістом, оптимізація сайта), так і нелегальні (створення подвійних сторінок, одна з яких демонструється відвідувачу, а інша сканується роботами пошукової мережі). Останнім часом широкою популярністю також користується оптимізація під соціальні медіа-ресурси SMO (social media optimization) та маркетинг в соціальних медіа SMM (social media marketing). Ця технологія використовує елементи піару та розміщує їх на соціальних медіа-сайтах.

Теоретично кожен рекламодавець може самостійно проводити контекстну рекламну кампанію в Інтернеті, проте це не є простою задачею. Для цього необхідними є не лише знання маркетингу, але й уміння ефективно працювати із системами розміщення реклами, що можливо лише за наявності інформації про їх особливості та можливості. Тому зручно користуватися послугами рекламних агентств, задачею яких є підтримка та підвищення ефективності рекламної кампанії. Завдяки спеціальним знанням та досвіду працівники таких агентств можуть проводити кампанію на більш високому рівні. Рекламодавець платить агентству за ефективність та якість проведення рекламної кампанії (підвищення відвідуваності сайта, збільшення обсягів продажів тощо). Альтернативний шлях проведення рекламної кампанії з використанням рекламного агентства як проміжної ланки зображено штриховою лінією.

Якщо власник сайта бажає розміщувати на ньому рекламні повідомлення, йому необхідно створити відповідний обліковий запис (наприклад, аккаунт Google AdSense). Автоматизовані системи відповідних рекламних кампаній розміщують на сайті конкретні оголошення зі схожою тематикою (дотримання принципів контекстної реклами). Власник сайта отримує певну платню за кожен клік (перехід) по оголошенню. Йому забороняється явним чином закликати відвідувачів натискати на рекламні повідомлення. Легальним методом підвищення зацікавленості користувачів є покращення дизайну, наповнення сайта корисними матеріалами тощо. Власник сайта та рекламна компанія підтримують двосторонній зв'язок, що приносить прибутки обом сторонам.

Таким чином, відвідувач (людина, яка натиснула на рекламне посилання) може переглядати рекламні повідомлення серед результатів пошуку, на сайтах та в соціальних мережах, блогах (відповідні зв'язки зображено штрих-пунктирною лінією). При цьому на рекламному майданчику розміщується або рекламне повідомлення (контекстна, банерна та інші види реклами), або посилання на відповідний сайт (SMM), або сам сайт як результат пошуку (SEO).

Всі відвідувачі можуть бути поділені на такі підкласи:

– клієнт;
– представник рекламного агентства;
– конкурент рекламодавця;
– власник сайта;
– «незалежний зловмисник».

В табл. 1 вказано мету відвідування для кожного із вказаних підкласів.

Таблиця 1 – Підкласи відвідувачів та їх цілі

Підклас відвідувача	Ціль відвідування
клієнт	ознайомлення із рекламним продуктом
представник рекламного агентства	вплив на якість рекламної кампанії, що проводиться іншими рекламними агентствами
конкурент рекламодавця	негативний вплив на якість рекламної кампанії рекламодавця
власник сайта	власне збагачення
«незалежний зловмисник»	інше

Особливості поведінки окремих груп зловмисників

З наведеної моделі взаємозв'язків можна виділити чотири групи потенційних зловмисників, кожній з яких властиві свої цілі та характер поведінки:

1) конкурент рекламодавця;
2) власник сайта;
3) рекламне агентство;
4) «незалежний зловмисник».

Конкурент рекламодавця може склікувати оголошення для розтрачання рекламного бюджету останнього та отримання вищої позиції під час показу. Для досягнення останньої мети зловмиснику необхідно повністю вичерпати денний бюджет рекламодавця (сума, яку рекламодавець згоден витратити на рекламу щоденно), котрий визначає – яку кількість кліків необхідно зробити для того, щоб припинився показ оголошення (оновлення денного бюджету відбувається в 00:00 кожного дня). В даному випадку фаза «проби» використовується саме для визначення розміру денного бюджету.

Ще одним вагомим фактором є часовий проміжок демонстрації оголошення. Засоби проведення рекламних кампаній надають можливість самостійно налаштовувати цей параметр. Очевидно, що найбільш прибутковим є той час, коли більшість потенційних клієнтів знаходяться в мережі. Для переважної більшості рекламних кампаній – це робочий день, тобто часові проміжки з 9:00 до 13:00 та з 14:00 до 18:00. Тому для досягнення успіху зловмисник має склікати оголошення або до початку робочого дня, або під час обідньої перерви (для фази «проби» ця вимога не є обов'язковою).

Під час проведення пробного та кінцевого нападів зловмиснику необхідно відключити власні рекламні кампанії, в іншому випадку він нашкодить самому собі. Позиція оголошення під час показу визначається двома параметрами: максимальною ставкою за клік – найбільша ціна, яку рекламодавець згоден платити за один перехід по оголошенню – та рейтингом кліків CTR (click through rate) – відношення кількості кліків до кількості показів впродовж деякого проміжку часу. Таким чином, генеруючи штучні кліки для оголошень конкурента, зловмисник генерує штучні покази для свого оголошення, що призводить до зниження рейтингу кліків та, відповідно, позиції оголошення.

Власник сайта може використовувати мережне шахрайство для власного збагачення, адже він отримує плату за кліки. На одному сайті розміщується велика кількість оголошень, отже, дії такого зловмисника будуть розосередженими по всіх наявних рекламних повідомленнях. Найважливішим для власника сайта є обережність, оскільки в даному випадку визначення конкретного зловмисника за наявності підозри є тривіальною задачею. Тому фаза «проба» необхідна для визначення особливостей реакції захисних систем рекламної кампанії на можливий напад (власник сайта може в будь-який момент переглянути стан свого бюджету, чи була отримана плата за кліки). «Атака» проводитиметься лише тоді, коли зловмисник буде впевненим у своїх діях. Досяг-

нення деякого високого рівня кліків не є обов'язковим. Навіть декілька кліків щоденно будуть приносити прибуток.

Для визначення недійсних кліків необхідно проводити аналіз трафіку всіх рекламних оголошень, що розміщені на сайті.

Рекламне агентство може використовувати склікування (споказування) як конкурентну боротьбу. Зменшення ефективності рекламних кампаній, що проводяться іншими рекламними агентствами, призводить до невдоволення рекламодавців і готує підґрунтя для переманювання клієнтів. Якщо рекламні кампанії рекламних агентств використовують однакові ключові слова, то необхідно відключити демонстрацію рекламних повідомлень кампаній, що проводяться зловмисником, аби не нашкодити власним клієнтам. При цьому під час проведення склікування зловмисник не має інших засобів перевірки ефективності своїх дій, окрім як припинення показу оголошення. Тому задачею і фази «проба», і фази «атака» є вичерпування денного бюджету оголошень. Це означає, що при проведенні розвідувального етапу кліки будуть більш розосередженими впродовж дня.

Для визначення наявності шахрайства, так само як і в попередньому випадку, логічно сумісно аналізувати показники всіх рекламних кампаній, що підтримуються даним рекламним агентством.

Останню групу складають «незалежні» зловмисники. Метою їх діяльності є підрив іміджу рекламної кампанії. Очевидно, що при виявленні наявності мережного шахрайства клієнти будуть незадоволені якістю захисних систем компанії та можуть припинити демонстрацію своїх оголошень або подати судовий позов. Для нанесення більшої шкоди обираються найдорожчі рекламні повідомлення, для яких ціна за один клік може досягати десятків доларів США.

Метою діяльності цієї групи не є пряме чи опосередковане самозбагачення, тому головною задачею є проведення нападу таким чином, щоб захисні системи не змогли відфільтрувати недійсні кліки. Фаза «проби» буде поступовою і, можливо, довготривалою (проводитиметься в декілька етапів).

Для двох останніх груп ефективним методом досягнення поставленої мети є споказування (зниження значення CTR призведе до зменшення якості рекламної кампанії).

Використання недиадного вейвлет-перетворення для визначення наявності аномальної поведінки користувачів

Зазначені вище особливості поведінки зловмисників проявляються у вигляді деяких «зайвих» екстремумів в числових рядах кількості кліків. Для виявлення таких непланових піків пропонується використовувати вейвлет-перетворення. Методи теорії вейвлетів на даний час застосовуються для пошуку особливостей заданого типу в сигналах різноманітної природи. Завдяки вейвлет-аналізу можна отримати не тільки інформацію про наявність в досліджуваному сигналі певної частоти (в даному випадку – кількості кліків), але й визначити моменти часу, в які відбувається зміна частот [11], [12].

Виділяють два класи вейвлет-перетворень – неперервні та дискретні. Незважаючи на те, що неперервне вейвлет-перетворення дозволяє отримати результати аналізу даних у зручному для подальшої інтерпретації вигляді, в практичних задачах найчастіше використовується дискретне вейвлет-перетворення, оскільки його можна обчислити за допомогою швидкого пірамідального алгоритму Малла.

В силу більш ефективної програмної реалізації зазвичай застосовується так зване диадне вейвлет-перетворення, тобто вейвлет-перетворення із коефіцієнтом масштабуван-

ня, який дорівнює двом. Однак не завжди його використання буде оптимальним, особливо коли шукані важливі особливості даних потрапляють в декілька сусідніх частотних інтервалів. В цих ситуаціях набагато ефективніше розділити частотну область на інтервали таким чином, щоб кожен частотний інтервал повністю охоплював особливості, які досліджуються. Подібну операцію можна виконати, обираючи відмінний від двох коефіцієнт масштабування, тобто використовуючи недиадні вейвлет-перетворення.

У випадку аналізу кількості кліків для виявлення зловмисників застосування диадного вейвлет-перетворення може не привести до бажаних результатів саме через те, що на кожному кроці диадного вейвлет-перетворення часовий діапазон кожного разу поділяється навпіл. При цьому після виконання двох кроків будуть отримані часові інтервали 0-6, 6-12, 12-18, 18-24, а після третього виконання вейвлет-розвинення – інтервали 0-3, 3-6, 6-9, 9-12, 12-15, 15-18, 18-21, 21-24. Таким чином, якщо зловмисник намагається провести атаку до початку робочого дня (наприклад, в проміжку між 4 та 8 годинами ранку), то після виконання як двох кроків вейвлет-розвинення, так і трьох кроків проміжок часу, який нас цікавить, потрапляє в два сусідні інтервали. Тому використання недиадних вейвлет-перетворень є більш оптимальним.

Авторами було проведено аналіз існуючих методів недиадного дискретного вейвлет-перетворення, виділені їх основні характеристики та надані рекомендації щодо класів задач, в яких може бути застосований той чи інший метод [13].

При часовому аналізі даних про кількість кліків має сенс розбиття повної доби на три восьмигодинних проміжки з подальшим розбиттям кожного з них на чотири двочасових (в останньому випадку двічі послідовно потрібно застосувати диадне перетворення), тобто потрібно використовувати вейвлет-перетворення із динамічною зміною коефіцієнта масштабування.

Найбільш оптимальним в такому випадку є застосування методу Поллока [14].

Висновки

У статті розглянуто ринок контекстної реклами, побудовано модель взаємодії учасників процесу проведення рекламної кампанії в мережі Інтернет.

На основі аналізу отриманої моделі були виділені чотири групи можливих зловмисників (людей, для яких мережне шахрайство може бути вигідним). Для кожної групи вказані цілі, що їй властиві, та особливості поведінки під час проведення нападу. Вони визначають відмінності шаблонів поведінки окремих груп від загального шаблону інформаційної атаки. В результаті були збудовані конкретні шаблони поведінки кожної групи та надані рекомендації щодо їх використання.

Визначення наявності аномалій, які відповідають шаблонам, пропонується виконувати за допомогою недиадного вейвлет-перетворення.

Література

1. Український ринок інтернет-реклами: прогнози експертів. – URL : http://video.management.com.ua/2010/06/03/ukrajinskyj-rynok-internet-reklamy-prohnozy-ekspertiv/
2. Украинский рынок интернет-рекламы: итоги 2007 года. – URL : http://ace.kiev.ua/ukrainskij-rynok-internet-reklamy-itogi-2007-goda/
3. Прогнозы объема рынка интернет-рекламы 2009 года. – URL : http://www.mediabusiness.com.ua/conent/view/-16387/48/lang,ru/
4. Украина: в 2009 году рынок интернет-рекламы впервые обогнал рынок рекламы на радио. – URL : http://www.seoteam.com.ua/article/internet-reklama/v-ukraine-rinok-internet-reklami-vpervie-obognal-rinok-reklami-na-radio.html

5. Рынок интернет-рекламы (поисковая оптимизация, контекстная и медийная реклама, веб-дизайн), емкость рынка и прогнозы развития. – URL : http://www.weblider.ru/article/article_97.html
6. Интернет займет 12,6% мирового рынка рекламы. – URL : http://reklamonstr.com/archives/4769
7. Matin S. Click Ahoy! Navigation Online Advertizing in a Sea of Fraudulent Clicks / S. Matin // Berkeley Technology Low Journal, Annual Review. – 2007. – V. 22, № 1. – P. 533-554.
8. "How does Google detects invalid clicks?". – URL : http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=6114
9. Goodman J. Pay-per-percentage of impressions: an advertising method that is highly robust to fraud / J. Goodman // Workshop on Sponsored Search Auctions. – 2005.
10. Asdemir K., Yahua M.A. Legal and strategic Perspectives on Click Measurement / K. Asdemir, M.A. Yahua // SEMPO Institute Opinions and Editorials. – 2006. – P. 11.
11. Малла С. Вейвлеты в обработке сигналов / Малла С. – М. : Мир, 2005. – 671 с.
12. Добеши И. Десять лекций по вейвлетам / Добеши И. – Ижевск : НИЦ «Регулярная и хаотическая динамика», 2001. – 464 с.
13. Чертов О.Р. Недиадні вейвлет-перетворення: дискретний випадок / О.Р. Чертов, В.В. Мальчиков // Труды научно-технической конференции с международным участием «Компьютерное моделирование в наукоемких технологиях» (КМНТ-2010), (18 – 21 мая 2010 г., г. Харьков). Часть 2. – Харьков, 2010. – С. 277-280.
14. Pollock D.S.G. Non-dyadic wavelet analysis / D.S.G. Pollock, I.L. Cascio // Optimisation, Econometric and Financial Analysis: Advances in Computational Management Science / E.J. Kontoghiorghes and C. Gatu (eds.) – Springer Verlag. – 2007. – Vol. 9. – P. 167-204.

О.Р. Чертов, Д.Г. Павлов, В.В. Мальчиков, М.В. Александрова

Выявление аномального поведения пользователя системы контекстной рекламы

В статье рассматривается рынок интернет-рекламы, в частности контекстной, формализуется взаимодействие объектов и субъектов процесса проведения рекламной кампании в сети. В результате анализа взаимосвязей строятся модели поведения возможных злоумышленников в системе, а также демонстрируется возможность применения недиадного вейвлет-преобразования для определения наличия аномального поведения пользователей.

O.R. Chertov, D.G. Pavlov, V.V. Malchykov, M.V. Alexandrova

Detection of the Abnormal Contextual Advertising System User Behavior

In current paper Internet-advertizing market, contextual advertising in particular, is considered; interaction between subjects and objects of the Internet-advertizing process is formalized. Relationship analysis results in possible fraudsters' behavior models construction. Nondiadic wavelet-transform use for the users' abnormal behavior detection is demonstrated.

Стаття надійшла до редакції 01.07.2010.